Uncertainty Quantification of GEOS-5 L-Band Radiative Transfer Model Parameters using Bayesian Inference and SMOS Observations

Gabriëlle J. M. De Lannoy[a,b], Rolf H. Reichle[a], Jasper A. Vrugt[c,d,e]

[a] *NASA Goddard Space Flight Center, Code 610.1, Greenbelt, MD 20771, USA*
[b] *GESTAR, Universities Space Research Association, Columbia, MD 21044, USA*
[c] *University of California, Irvine, Department of Civil and Environmental Engineering, Irvine, CA 92697-2175, USA*
[d] *University of California, Irvine, Department of Earth System Science, Irvine, CA 92697-2175, USA*
[e] *Institute for Biodiversity and Ecosystems Dynamics, University of Amsterdam, 1098 XH Amsterdam, The Netherlands*

Abstract

Uncertainties in L-band (1.4 GHz) microwave radiative transfer modeling (RTM) affect the simulation of brightness temperatures (Tb) over land and the inversion of satellite-observed Tb into soil moisture retrievals. In particular, accurate estimates of the microwave soil roughness, vegetation optical depth and scattering albedo for large-scale applications are difficult to obtain from field studies and often lack an estimate of uncertainty. Here, a Markov Chain Monte Carlo (MCMC) simulation method is used to determine satellite-scale estimates of RTM parameters and their posterior uncertainty by minimizing the misfit between long-term averages and standard deviations of simulated and observed Tb at multiple incidence angles, at horizontal and vertical polarizations, and for morning and evening overpasses. Tb simulations are generated with the land model component of the Goddard Earth Observing System (version 5) and confronted with Tb observations from

the Soil Moisture Ocean Salinity (SMOS) satellite mission. The maximum a posteriori density (MAP) parameter values reduce the root-mean-square differences between observed and simulated long-term Tb averages and standard deviations to 3.4 K and 2.3 K, respectively. The relative uncertainty of the posterior RTM parameter estimates is typically less than 25% of the MAP parameter value, whereas it exceeds 100% for literature-based prior parameter estimates. It is also shown that the parameter values estimated through Particle Swarm Optimization are in close agreement with those obtained from MCMC simulation. The MCMC results for the RTM parameter values and the uncertainties presented herein are directly relevant to the need for accurate Tb modeling in global land data assimilation systems.

Keywords:

radiative transfer modeling, brightness temperature, Bayesian parameter estimation, uncertainty, Markov Chain Monte Carlo simulation, SMOS

1. Introduction

Uncertainties in radiative transfer modeling (RTM) affect the simulation of brightness temperatures (Tb) over land and the inversion of satellite-observed Tb to soil moisture retrievals. Quantification of these uncertainties is crucial to producing, validating and using L-band (1.4 GHz) passive microwave data, such as those obtained from the Soil Moisture Ocean Salinity (SMOS, Kerr et al. (2010)) and future Soil Moisture Active Passive (SMAP, Entekhabi et al. (2010)) missions. Yet, it is not particularly clear which RTM formulation and parameter values to use for large-scale applications.

In the context of large-scale forward Tb simulation, several studies have analyzed the effect of different RTM formulations for the microwave roughness length, vegetation parameterization and soil dielectric model (Drusch et al., 2009; de Rosnay et al., 2009), the impact of parameter values (De Lannoy et al., 2013) and the sensitivity to dynamic land surface variables (Balsamo et al., 2006). Similarly, soil moisture retrievals based on Tb observations are affected by the RTM formulation and parameter values (Crosson et al., 2005; Panciera et al., 2009; Konings et al., 2011; Parinussa et al., 2011), as well as by the choice of background and auxiliary fields, such as soil temperature and vegetation characteristics (Kerr et al., 2012; O'Neill et al., 2012). Collectively, these studies suggest that RTMs exhibit significant uncertainty and that the impact of this uncertainty on large-scale Tb simulations and soil moisture retrievals remains unclear.

Estimating microwave RTM parameters and their uncertainty is a major challenge, especially at larger spatial scales. Field experiments have provided RTM parameter values (de Rosnay et al., 2006; Grant et al., 2007; Panciera

et al., 2009; Sabater et al., 2011), but mostly without an underlying estimate of their uncertainty. De Lannoy et al. (2013) derived global-scale RTM parameter values and ad hoc uncertainty estimates using SMOS observations and Particle Swarm Optimization (PSO, Kennedy and Eberhart (1995)). PSO is specifically designed to find the optimal parameter values, without recourse to estimating their underlying uncertainty.

In this paper, we apply a (Bayesian) Markov chain Monte Carlo (MCMC) simulation method to estimate the posterior RTM parameter distribution. The DiffeRential Evolution Adaptive Metropolis ($\text{DREAM}_{(\text{ZS})}$) algorithm is used with parallel direction and snooker sampling from past states (Vrugt et al., 2008, 2009; Laloy and Vrugt, 2012). Bayesian approaches such as $\text{DREAM}_{(\text{ZS})}$ have several advantages over optimization methods such as PSO. The explicit treatment and analysis of uncertainty help to understand which parts of the RTM are well resolved and which elements require further attention. Furthermore, a formal analysis of the residuals can be used to check the validity of our assumptions about the probabilistic properties of the errors and to discern whether reliable parameter values have been derived.

Implementation of the Bayesian paradigm coupled with MCMC simulation comes at an increased computational expense. Full exploration of the posterior distribution is very costly, and hence difficult to warrant in global scale operational applications that rely on evolving modeling systems in need of frequent re-calibrations. Yet, this paper will demonstrate that the proposed approach provides important insights into the uncertainty of large-scale RTM parameters, and can be used to benchmark the results from optimization algorithms such as PSO.

The goals of the present paper are thus to infer large-scale RTM parameters and their posterior uncertainty using a Bayesian method, and to study the associated simulated Tb uncertainty. This research complements and advances the work of De Lannoy et al. (2013) where PSO was used to find parameters without formal statistical estimates of uncertainty. We are using the Goddard Earth Observing System, version 5 (GEOS-5) modeling framework that will be used to generate the planned global SMAP Level 4 Surface and Root Zone Soil Moisture (L4_SM) data product through assimilation of SMAP Tb observations (Reichle et al., 2012). Here, the focus is on estimating time-invariant RTM parameters and their uncertainty by minimizing climatological (long-term) differences between multi-angular, horizontally and vertically polarized Tb for morning and evening overpasses from SMOS observations and GEOS-5 simulations, while explicitly treating and estimating their respective uncertainties.

The time-invariant estimated parameters will be used in a data assimilation system (outside the scope of this paper), where state variables such as soil moisture and soil temperature will be updated in response to short-term variations in the observed Tb. Residual long-term Tb errors, or biases, that remain after the estimation of the RTM parameters will be addressed through model refinement and within the data assimilation system. The (Bayesian) uncertainties of the RTM parameters and the corresponding estimates of the residual observation and model error presented here will help with the development of such data assimilation systems. Furthermore, the uncertainties of the RTM parameters reported here will also facilitate sensitivity analyses of soil moisture retrieval algorithms (Parinussa et al., 2011).

To summarize, in this paper we apply MCMC simulation using multi-angular SMOS Tb observations to (i) determine if the maximum a posteriori density (MAP) parameter values derived from the posterior distribution sampled with DREAM$_{(ZS)}$ can be approximated using PSO, (ii) obtain reliable estimates of parameter uncertainty, and (iii) quantify the impact of errors in parameters and other sources on Tb simulations. The remainder of this paper is organized as follows. Section 2 summarizes the modeling system and the SMOS observations used in the present study. This is followed in section 3 by a brief description of the DREAM$_{(ZS)}$ MCMC simulation method. This section also discusses several quantitative diagnostic metrics to analyze the simulated Tb uncertainty. Section 4 discusses the main findings and results of this paper. This is followed in section 5 with a summary and conclusions.

2. Observations and Model

2.1. SMOS Tb Data

Since its launch in November 2009, the SMOS mission provides global L-band Tb data at a nominal spatial resolution of 43 km. On average, a given location on the equator is revisited once every 3 days. Here we use the multi-angular, full polarization Tb data from the period 1 July 2010 to 1 July 2012. Specifically, the data are extracted from the MIR_SCLF1C product, with processor version 504 for the years 2010 and 2011 (reprocessed in 2012), and version 551 from January 2012 onwards. De Lannoy et al. (2013) discuss in detail the various steps involved in the processing of the SMOS data. Most importantly, the data are screened extensively using both product-based data quality information and model-based quality control rules. Furthermore, the

data are spatially mapped onto a 36 km Equal-Area Scalable Earth (EASE) grid and binned per incidence angle. Consistent with De Lannoy et al. (2013), only a subset of 6 incidence angles is used: θ=[32.5°, 37.5°, 42.5°, 47.5°, 52.5°, and 57.5°], where, for example, 32.5° represents the average of all Tb data with incidence angles between 32 and 33 degrees.

To estimate the microwave RTM parameters, long-term averages ($\boldsymbol{m}_o$) and standard deviations ($\boldsymbol{s}_o$) of the SMOS data are computed separately for each of the 6 incidence angles, 2 polarizations (horizontal H and vertical V), and 2 overpass times (ascending at ∼06:00h local time (LT), descending at ∼18:00h LT). This results in a total of 48 "observations" per grid cell: 24 ($= 6 \times 2 \times 2$) observations of the long-term average Tb and 24 observations of the long-term Tb standard deviation. Section 3 provides a more extensive description of how these 48 observations are used.

2.2. GEOS-5 Tb Modeling

The modeling combines (i) land surface modeling with the Catchment land surface model (CLSM) and (ii) radiative transfer modeling with a tau-omega model to simulate long-term Tb averages and standard deviations. As in De Lannoy et al. (2013), the GEOS-5 CLSM (Koster et al., 2000) is set up on the 36 km EASE grid and spun up prior to the SMOS observation period. Surface meteorological forcing data at a 1/2°×2/3° spatial and hourly temporal resolution are taken from the Modern-Era Retrospective analysis for Research and Applications (MERRA, Rienecker et al. (2011)). The MERRA-precipitation is corrected with gauge-based precipitation from the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center "Unified" (CPCU) product (Reichle, 2012). The model version is the

same as that used for the MERRA-Land data product (Reichle et al., 2011), except for two changes that more closely align the model with the version that will ultimately be used for the SMAP L4_SM data product: (i) the surface soil moisture pertains to the top 5 cm surface layer (as opposed to the top 2 cm layer in MERRA-Land), and (ii) a preliminary version of updated soil parameters from a forthcoming version of GEOS-5 is used.

The vegetation parameterization in CLSM uses 8 default vegetation classes. For the RTM simulations, these classes are further refined into the 16 classes defined by the Moderate Resolution Imaging Spectroradiometer (500 m MOD12Q1 V004) International Geosphere-Biosphere Programme (IGBP) land cover classification (Loveland and Belward, 1997). Figure 1 shows the North American study domain which includes 9 of the 16 IGBP vegetation classes.

In essence, Tb is determined by the surface soil temperature, soil moisture and other soil and vegetation characteristics. To simulate L-band Tb, the prognostic CLSM soil moisture, soil temperature, vegetation water content, air temperature and climatological vegetation dynamics are used as inputs to a diagnostic zero-order (tau-omega) microwave RTM, briefly described in Appendix A. The key model parameters that determine the microwave surface roughness h, the scattering albedo ω, and vegetation optical depth τ will be estimated using the multi-angular SMOS observations (section 3). As outlined in Appendix A, h is a function of soil moisture and the time-invariant parameters h_{min} and h_{max} (Eq. A.4), and τ depends on the leaf area index (LAI) and the time-invariant parameters b_H and b_V (Eq. A.6).

3. Methods

3.1. Overview

Keeping up with De Lannoy et al. (2013), the objective of the parameter estimation is to minimize the differences in long-term (climatological) averages and standard deviations between multiple types of SMOS-observed and GEOS-5-modeled Tb. We purposely do *not* minimize differences in the time domain as the goal of the present paper is to derive parameter estimates that result in the smallest possible long-term bias in the simulation of Tb. Short-term differences between Tb observations and simulations will be addressed in future studies using sequential data assimilation. We estimate a time-invariant multi-dimensional parameter set, hereafter referred to as $\boldsymbol{\alpha}$, that determines climatological features of the simulated Tb. The parameters are estimated locally, i.e., for each grid cell independently, and only for non-frozen land surface conditions as determined by the GEOS-5 modeling system.

In addition to finding parameter estimates that result in accurate Tb simulations, the Bayesian methodology used here samples the complete posterior parameter distribution. This distribution summarizes the dispersion of the parameters, and can be used to derive the model simulation uncertainty by propagating each sample of the posterior through the model. If our probabilistic assumptions about the constituent error sources that affect the Tb simulations are correct, then our analysis will provide maximum a posteriori density (MAP) parameter estimates that accurately reflect their true values, and the underlying posterior uncertainty will be statistically meaningful. In other words, we seek accurate and precise parameter values with a minimum

simulation bias, and statistically meaningful spread.

Table 1 gives an overview of the parameters estimated in different experiment cases. All cases estimate the 5 most relevant RTM parameters: h_{min}, $\Delta h \equiv h_{max} - h_{min}$, b_H, $\Delta b \equiv b_V - b_H$ and ω. Simultaneous inference of these select parameters accounts for their correlation in minimizing the error residuals, and such an approach is thus preferred over sequential (stepwise) fitting of the individual parameters (De Lannoy et al., 2013). Based on these *time-invariant* parameters, *time-variant* values of h, τ_H and τ_V are computed, using dynamic information about soil moisture for h (Eq. A.4) and LAI for τ (Eq. A.6). Time-averaged results for $< h >$ and $< \tau >$ are then presented, where $< \cdot >$ denotes the long-term time average. The RTM parameters are estimated with DREAM$_{(ZS)}$ (section 3.2) or PSO (Appendix B), hereafter referred to as cases D and P, respectively, both of which estimate h_{min}, Δh, b_H, Δb and ω. Furthermore, a third case, D$_\sigma$, additionally estimates the residual Tb error statistics σ_m and σ_s, using DREAM$_{(ZS)}$ (discussed below). We thus estimate 5 parameters per grid cell for cases P and D, and 7 parameters per grid cell for case D$_\sigma$.

To derive these parameters, we minimize per grid cell the *climatological*, or long-term, differences between 48 Tb observations and simulations. These $2 \times 24 = 48$ observations consist of long-term Tb averages and Tb standard deviations for the 24 combinations of 2 polarizations, 2 overpass times, and 6 incidence angles. For simplicity, the errors in these observations are assumed to be independent, that is, we neglect correlations in instrument errors and errors between H- and V-polarized observations at identical incidence angles. Similarly, the simulation errors are assumed to be independent, even though

some correlation is to be expected. Cross-correlations in observation or simulation errors between various incidence angles would increase the uncertainty in the posterior parameter estimates. Note that temporal correlations in the errors are of little concern because the observations are long-term averages and standard deviations, and not measurements in the time domain (Wöhling and Vrugt, 2011).

In keeping up with De Lannoy et al. (2013), the two years of historical SMOS data are divided into a calibration period (1 July 2011 - 1 July 2012) and an evaluation period (1 July 2010 - 1 July 2011). To ensure a meaningful calibration at each grid cell, we impose a minimum of 20 valid data points (N_i) per year and per overpass time to compute the long-term Tb average and standard deviation for a particular combination ($i = 1, \ldots, 24$) of incidence angle, polarization and overpass time. The requirement of $N_i \geq 20$ is used for the calculation of evaluation statistics as well. In addition, for the calibration, we always require a minimum total number of data points of 480=20×6 angles×2 polarizations×2 overpass times.

This study relies on proper cross-validation using independent calibration and evaluation periods. Preparatory work for the forthcoming operational SMAP L4_SM product uses all available historical SMOS data in the calibration, and the results are comparable to what is reported herein. Given the similar nature of L-band observations from SMOS and SMAP, we expect that parameters calibrated with SMOS observations will initially serve well in the SMAP L4_SM system, while a sufficient SMAP data record is being accumulated for later re-calibration.

3.2. Markov Chain Monte Carlo (MCMC) Sampling

The Bayesian paradigm provides a framework for the treatment of all sources of uncertainty in modeling Tb. In this paper, we focus on parameter uncertainty, and treat the other sources of error as a single lumped term (details to follow). The posterior probability distribution of the parameters is computed by combining the observation likelihood $p(\boldsymbol{m}_o, \boldsymbol{s}_o|\boldsymbol{\alpha})$ with a prior distribution $p(\boldsymbol{\alpha})$:

$$p(\boldsymbol{\alpha}|\boldsymbol{m}_o, \boldsymbol{s}_o) = \frac{p(\boldsymbol{m}_o, \boldsymbol{s}_o|\boldsymbol{\alpha})p(\boldsymbol{\alpha})}{\int_{\boldsymbol{\alpha}} p(\boldsymbol{m}_o, \boldsymbol{s}_o|\boldsymbol{\alpha})d\boldsymbol{\alpha}} \tag{1}$$

The observations consist of long-term averages ($m_{i,o} \in \boldsymbol{m}_o$) and standard deviations ($s_{i,o} \in \boldsymbol{s}_o$) of Tb for 24 different combinations of incidence angles, polarizations and overpass times ($i = 1, \ldots, 24$). The denominator is a normalization factor and thus it suffices to maximize $p(\boldsymbol{m}_o, \boldsymbol{s}_o|\boldsymbol{\alpha})p(\boldsymbol{\alpha})$ to find the posterior distribution of $\boldsymbol{\alpha}$. In practice, it is difficult to solve this problem analytically and we therefore resort to MCMC simulation to generate a sample of the posterior target distribution.

In this paper, the DREAM$_{(\mathrm{ZS})}$ algorithm (Vrugt et al., 2008; Laloy and Vrugt, 2012) with sampling from past states is used to efficiently explore the posterior parameter distribution. This algorithm adaptively updates the scale and orientation of the proposal distribution during sampling, and is specifically designed to rapidly explore multi-dimensional target distributions. In DREAM$_{(\mathrm{ZS})}$, multiple chains are running in parallel and the update of a chain is determined from an external sample of points that collectively summarizes the search history of all the individual chains. The log-likelihood of the current and proposed parameter values are compared

using the Metropolis selection rule. If the proposal is accepted, the chain moves to this new point, otherwise the chain remains at its current position. Diminishing adaptation of the external archive of samples ensures convergence to the exact posterior distribution.

We assume a Gaussian prior for each of the individual parameters $\alpha_{0,k} \in \boldsymbol{\alpha}_0$. The mean and standard deviation of this multi-normal distribution $p(\boldsymbol{\alpha})$ are derived from literature values summarized in Table 1. Note that these values were referenced as 'Lit2' in De Lannoy et al. (2013). The prior RTM parameters were subjectively selected out of a range of possible values in the literature. The prior mean of parameter k is given by a (vegetation-dependent) value $\alpha_{0,k}$ and the standard deviation $\sigma_{\alpha_{0,k}}$ is computed as $\sigma^2_{\alpha_{0,k}} = (\alpha_{max,k} - \alpha_{min,k})^2/12$, where $\alpha_{max,k}$ and $\alpha_{min,k}$ denote the respective upper and lower bounds.

The following log-likelihood function is used to minimize the differences in long-term Tb averages and standard deviations between the observations $(m_{i,o}, s_{i,o})$ and corresponding simulations $(m_i(\boldsymbol{\alpha}), s_i(\boldsymbol{\alpha}))$:

$$L = \ln(p(\boldsymbol{m}_o, \boldsymbol{s}_o|\boldsymbol{\alpha})) = -\frac{24}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{24}\ln(\sigma^2_{i,m}) - \sum_{i=1}^{24}\frac{(m_{i,o} - m_i(\boldsymbol{\alpha}))^2}{2\sigma^2_{i,m}}\Bigg\} L_{m,o}$$
$$- \frac{24}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{24}\ln(\sigma^2_{i,s}) - \sum_{i=1}^{24}\frac{(s_{i,o} - s_i(\boldsymbol{\alpha}))^2}{2\sigma^2_{i,s}}\Bigg\} L_{s,o} \qquad (2)$$

This formulation thus explicitly takes into consideration long-term biases in the Tb average ($L_{m,o}$ [-]) and the Tb variability ($L_{s,o}$ [-]) and is derived from

a classical Gaussian likelihood function:

$$p(\boldsymbol{m}_o, \boldsymbol{s}_o|\boldsymbol{\alpha}) = \prod_{i=1}^{24} \left[\frac{1}{\sqrt{2\pi\sigma_{i,m}^2}} \exp\left(-\frac{(m_{i,o} - m_i(\boldsymbol{\alpha}))^2}{2\sigma_{i,m}^2} \right) \right] \cdot \prod_{i=1}^{24} \left[\frac{1}{\sqrt{2\pi\sigma_{i,s}^2}} \exp\left(-\frac{(s_{i,o} - s_i(\boldsymbol{\alpha}))^2}{2\sigma_{i,s}^2} \right) \right] \quad (3)$$

where $\sigma_{i,m}$ and $\sigma_{i,s}$ denote the (ensemble) standard deviations of the residual differences between the observed and simulated values of the long-term Tb averages and standard deviations, respectively. These standard deviations provide a lumped description of model, input data, and observation errors.

3.3. Likelihood, Objective Function and Algorithm Settings

The design of the DREAM$_{(ZS)}$ likelihood function L (Eq. 2) and the PSO objective function J (Eq. B.1) warrants further discussion. As discussed above, we sample the climatological, or long-term, Tb averages and standard deviations over multiple incidence angles, polarizations and overpass times (that is, 2×24 observations, $i = 1, \ldots, 24$) per location, rather than a series of observations at multiple times. The long-term Tb averages and standard deviations could also be interpreted as 'summary statistics' or 'signatures' of the system, and hence our approach has elements in common with the diagnostic model evaluation procedure presented in Vrugt and Sadegh (2013).

The variables $\sigma_{i,m}$ and $\sigma_{i,s}$ measure the (ensemble) standard deviation of the residual differences between the observed and simulated long-term Tb averages and standard deviations, respectively, for each observation i. The residual errors are assumed to have a zero mean and include both SMOS observation and simulation errors, due to, e.g., inaccurate soil moisture, tem-

perature or vegetation characteristics. These $\sigma_{i,m}$ and $\sigma_{i,s}$ statistics trade off errors in the long-term Tb averages against those of the long-term Tb standard deviations (as well as deviations from the prior parameter constraints). Since only one sample is available for each observation, it is impossible to estimate individual $\sigma_{i,m}$- and $\sigma_{i,s}$-values. Therefore, $\sigma_{i,m}$ and $\sigma_{i,s}$ include a homoscedastic term (σ_m, σ_s) and a heteroscedastic factor w_i to account for the robustness of the diagnosed long-term Tb averages and standard deviations, i.e. $\sigma_{i,m}^2 = w_i \sigma_m^2$ and $\sigma_{i,s}^2 = w_i \sigma_s^2$. The homoscedastic term is identical for all 24 observations, and we either set σ_m and σ_s to a default value of 1 K (De Lannoy et al., 2013), or alternatively we estimate σ_m and σ_s jointly with the RTM parameters (see section 3.1). The weights are given by $w_i = \frac{N}{N_i}$, where N_i denotes the number of data points in time that contribute to a particular long-term Tb average (or standard deviation), and N signifies the average contributing number of time steps across all 24 observations. These weights are typically close to 1 and assign somewhat more (less) weight to climatological Tb differences that are based on more (fewer) individual data points in the different 1-year data time series that underlie the 24 different observations. For example, Tb observations at low incidence angles are by design based on fewer instantaneous data points than Tb observations at high incidence angles.

Per grid cell, a maximum of 12,000 log-likelihood function evaluations are performed with DREAM$_{(ZS)}$ using standard settings of the algorithmic variables. For PSO, we use the same algorithmic settings as reported in De Lannoy et al. (2013), except that a swarm size of 10 particles is used with a minimum of 10 and maximum of 100 iterations. The search is terminated

if the reduction of the objective function is smaller than 1E-5 over the last 10 iterations. A total of 12 repetitions are performed, which results in a maximum of 12,000 function evaluations.

3.4. Posterior Parameter Distribution

The MAP parameter values are defined as those with the largest value for L (Eq. 2, DREAM$_{(ZS)}$) or the smallest value for J (Eq. B.1, PSO). These MAP parameter estimates will subsequently be used in the RTM that is part of the Tb assimilation system for state updating (not discussed herein). Note that these MAP values are not necessarily identical to the posterior ensemble mean of the distribution. For the DREAM$_{(ZS)}$ experiments, the last 25% of the MCMC chains (3,000 samples) are used to summarize parameter uncertainty by calculating the standard deviation of each individual parameter. To illustrate this in more detail for one grid cell, consider Fig. 2a, which depicts the marginal distributions of the RTM parameters. We define the uncertainty as the ensemble standard deviation $stdv[\alpha] \equiv \sqrt{\overline{(\alpha - \overline{\alpha})^2}}$ centralized around the ensemble mean $\overline{\alpha}$, not around the MAP parameter value α_{MAP}. The notation $\overline{\cdot}$ refers to the ensemble mean. Note that the standard deviation around the MAP estimate $stdv_{MAP}[\alpha]$ can be found as a function of the centralized standard deviation $stdv[\alpha]$, i.e. $stdv_{MAP}[\alpha]^2 = stdv[\alpha]^2 + \Delta_\alpha^2$, where $\Delta_\alpha = \overline{\alpha} - \alpha_{MAP}$ is the difference between the ensemble mean and MAP parameter estimate. We found that, across the different experiments, Δ_α is usually small (see section 4.1), so that $stdv_{MAP}[\alpha] \sim stdv[\alpha]$.

3.5. Tb Simulation Performance and Uncertainty

A number of measures are used to evaluate the long-term Tb simulations and their associated uncertainty. Fig. 2b illustrates some of the terms used in this evaluation. We assess the quality of the deterministic Tb simulations with the MAP parameter estimates, using the mean-square difference (MSD [K^2]) between the observed and simulated long-term Tb averages (Eq. 4) and standard deviations (Eq. 5) across the 24 different observations:

$$MSD_m = \frac{1}{24}\sum_{i=1}^{24}(m_i(\boldsymbol{\alpha}_{MAP}) - m_{i,o})^2 \quad (4)$$

$$MSD_s = \frac{1}{24}\sum_{i=1}^{24}(s_i(\boldsymbol{\alpha}_{MAP}) - s_{i,o})^2 \quad (5)$$

If the modeling errors were solely due to inaccurate parameter values and the observational terms were error-free, these metrics should be very close to zero. In practice, however, the model structure is not perfect and model inputs such as soil moisture and temperature as well as the observational terms are subject to errors. Therefore, the metrics will substantially deviate from zero and reflect the total residual simulation and observation errors. The 24 differences contributing to MSD_m are illustrated as Δ_{m_i} in Fig. 2b.

If the uncertainties are well estimated and biases between observations and simulations are constrained during the calibration, then the "actual" (MSD_m, MSD_s) and "expected" ensemble ($\sigma^2_{i,m}$, $\sigma^2_{i,s}$) residual Tb error variances should be equal, or their ratio should be close to 1. Note that a similar check of consistency is used to verify the prescribed observation and simulation uncertainties in data assimilation systems (Reichle et al., 2002) and for ensemble forecast verification (De Lannoy et al., 2006). The only difference is

that here, the mean values (i.e. the 'M', or mean, in MSD) are derived from multiple observations types ($i = 1, \ldots, 24$), whereas in the earlier studies the mean was calculated in the time domain.

The above total residual Tb error lumps all long-term errors in the model, input data, and observations. The Tb simulation error due to parameter uncertainty can be isolated. To this end, we analyze an ensemble of Tb simulations, obtained by propagating 20 samples from the MCMC derived posterior parameter distributions through the RTM. The performance of the ensemble mean of simulated long-term Tb averages $\overline{m_i(\boldsymbol{\alpha})}$ and standard deviations $\overline{s_i(\boldsymbol{\alpha})}$ is given by:

$$MSD_{\overline{m}} = \frac{1}{24}\sum_{i=1}^{24}\left(\overline{m_i(\boldsymbol{\alpha})} - m_{i,o}\right)^2 \quad (6)$$

$$MSD_{\overline{s}} = \frac{1}{24}\sum_{i=1}^{24}\left(\overline{s_i(\boldsymbol{\alpha})} - s_{i,o}\right)^2 \quad (7)$$

where $\overline{\cdot}$ denotes the ensemble mean. Fig. 2b illustrates the 24 differences contributing to $MSD_{\overline{m}}$ as $\Delta_{\overline{m_i}}$. Because of the non-linear nature of the RTM and because $\overline{\boldsymbol{\alpha}}$ and $\boldsymbol{\alpha}_{MAP}$ generally differ, $MSD_{\overline{m}}$ and $MSD_{\overline{s}}$ will deviate from MSD_m and MSD_s, but the differences should be limited if the posterior RTM parameter distributions are narrow.

The Tb uncertainty due to parameter error is quantified by the ensemble standard deviation in simulated long-term Tb averages ($\sigma_{i,m,par}$, illustrated in Fig. 2b) and long-term Tb standard deviations ($\sigma_{i,s,par}$). The corresponding ensemble variance for each observation type i is:

$$\sigma^2_{i,m,par} = \overline{\left(m_i(\boldsymbol{\alpha}) - \overline{m_i(\boldsymbol{\alpha})}\right)^2} \quad (8)$$

$$\sigma^2_{i,s,par} = \overline{\left(s_i(\boldsymbol{\alpha}) - \overline{s_i(\boldsymbol{\alpha})}\right)^2} \quad (9)$$

Averaged across 24 observations, this results in a mean ensemble spread of:

$$MEnSp_{m,par} = \frac{1}{24}\sum_{i=1}^{24}\sigma^2_{i,m,par} \quad (10)$$

$$MEnSp_{s,par} = \frac{1}{24}\sum_{i=1}^{24}\sigma^2_{i,s,par} \quad (11)$$

The above metrics are primarily important to quantify the skill of the posterior Tb simulations. For comparison, we also calculate the MSD_m and MSD_s using prior RTM parameters (Table 1). For clarity, all metrics in this section are expressed as variances [K^2], but in the results below we present Tb uncertainties as the corresponding standard deviations [K].

4. Results

4.1. RTM Parameters and Their Uncertainty

In this section, we analyze the MAP values of the microwave surface roughness $< h >$, the vegetation optical depth $< \tau >$, the scattering albedo ω, and their posterior uncertainty ($stdv[.]$). As will be shown below, the $DREAM_{(ZS)}$ case D_σ should be considered as benchmark throughout the paper, because of the statistical rigor of the sampled posterior. Fig. 3 shows maps of the prior parameter values and the MAP estimates derived from experiment cases P, D and D_σ (Table 1). The spatially averaged posterior parameter values are very similar for all 3 cases, with a microwave roughness $< h >$ around 0.72±0.5 [-], a vegetation optical depth $< \tau >$ of 0.26±0.15 [-] and a scattering albedo ω of 0.09±0.07 [-], where the values after the ± sign measure the spatial standard deviation and reflect the variability of the MAP parameters across the spatial domain. Note that these latter values should

not be confused with uncertainty estimates. Compared to the prior values, $< h >$ is generally larger for grassland, $< \tau >$ is smaller for forests and ω is larger for all vegetation classes except grassland (details per vegetation class not shown; these finding are similar to those of De Lannoy et al. (2013)). The spatial patterns for the 3 experiments are also very similar, with exceptions discussed below. Moreover, Fig. 3 suggests that MAP values derived with the PSO algorithm closely match those of $\text{DREAM}_{(\text{ZS})}$.

Fig. 4 shows the ensemble parameter uncertainty for cases D and D_σ. Maps with RTM parameter uncertainty estimates for PSO (obtained as in De Lannoy et al. (2013)) are not shown, because they are statistically invalid and significantly larger than those derived with $\text{DREAM}_{(\text{ZS})}$. The prior parameter uncertainty is also not shown, because it is spatially uniform. Table 1 indicates that the relative uncertainty in the literature-based prior parameters exceeds 100% of $\boldsymbol{\alpha}_0$. In contrast, the relative uncertainties for case D (Fig. 4a-c) are less than 10% of the MAP parameter value and substantially smaller than the spatial variability in the MAP values. For case D_σ, the relative uncertainties are ranging up to 25% of the MAP values: for $< h >$, the spatially averaged uncertainty is 0.10 [-], for $< \tau >$ 0.04 [-] and for ω 0.02 [-]. The uncertainty in $< h >$ typically increases with more complex terrain and is smallest in the cropped region southwest of the Great Lakes. The uncertainty of $< \tau >$ is largest in the forested Appalachian mountains where the highest MAP values of $< \tau >$ are found. On the contrary, ω is best defined in this area and uncertainties in ω increase in the dry Western mountain ranges. The $< h >$-values exhibit more uncertainty where either the uncertainty in ω (Fig. 4e) or $< \tau >$ (Fig. 4f) is larger. The global mean

absolute differences between the MAP and ensemble mean parameter values (not shown; discussed in section 3.4) for $< h >$, $< \tau >$ and ω are $\Delta\alpha$=0.07, 0.02 and 0.01 [-], respectively, for case D_σ, and $\Delta\alpha$=0.02, 0.01 and 0.00 [-], respectively, for case D.

In summary, both $DREAM_{(ZS)}$ cases D and D_σ provide MAP parameter values that are very similar and in close agreement with the PSO estimates (Fig. 3). The $DREAM_{(ZS)}$ derived posterior parameters appear well defined with relative uncertainties that are less than 25% of the MAP values. It will be shown below that the uncertainty estimates of case D_σ – unlike those of case D – are consistent with the sample root-mean-square difference (RMSD) between long-term Tb observations and simulations.

4.2. Residual Tb Error Standard Deviations

To analyze the differences between cases D and D_σ, or specifically the effect of estimating the residual error standard deviations of the long-term Tb averages (σ_m) and standard deviations (σ_s), we list the domain-averaged MAP parameter values and their associated uncertainties in Table 2 for all experiment cases. In addition, Fig. 5 depicts the results for different vegetation classes. As discussed above, cases D and D_σ return similar MAP values for the RTM parameters with some local exceptions, such as for example for ω over cropland (Fig. 5e). The estimated posterior RTM parameter uncertainty increases about 2 - 3 times, when σ_m and σ_s are included in the parameter estimation (i.e., case D_σ). For case D_σ, the domain-averaged values are $\sigma_m = 3.5$ K and $\sigma_s = 2.3$ K (Table 2), whereas case D uses default values of $\sigma_m = \sigma_s$=1 K. The σ_m and σ_s estimates are not defined for the simulations with prior parameters.

The residual error standard deviations σ_m and σ_s are estimated to be larger than the initially imposed 1 K because of significant Tb observation and simulation errors that are not due to parameter errors. It is necessary to use this increased residual Tb error in the RTM parameter estimation to ensure parameters that do not, or at least only minimally, compensate for errors other than parameter error (i.e. error in geophysical input fields, auxiliary information, RTM structure or in observations). The local differences between the MAP values for cases D and D_σ are thus explained by how much the parameter values are forced to compensate for errors other than parameter error. As will be shown below, this can result in suboptimal parameter values for case D. The uncertainty in the RTM parameters for D_σ is larger, because we allow for more realistic Tb simulation and observation error, resulting in a wider range of acceptable RTM parameters. From a Bayesian perspective, the observation likelihood function becomes wider (larger σ_m and σ_s) and thus the posterior parameter probability function is expected to be wider.

The value of σ_m and its posterior uncertainty are largest in cropped regions (Fig. 5g) where residual Tb errors are dominated by less skillful model simulations. This is to be expected because irrigation is not simulated and the climatological LAI estimates do not account for interannual crop rotations. The parameters should not compensate for these errors, and the default values of $\sigma_m = \sigma_s = 1$ K make cases D and P vulnerable to suboptimal solutions. For example, the relatively larger differences between D and D_σ for σ_m over cropland result in larger differences in their respective MAP values of ω (Fig. 5i). For forests, $\sigma_s = 1$ K appears to be a good estimate

(Fig. 5i) because the variability in Tb is expected to be low due to vegetation attenuation.

Both the MAP values and the uncertainties for σ_m are always larger than those derived for σ_s. One of the reasons for the higher σ_m values are the opposite signs in the biases for the long-term Tb averages at morning and evening overpasses, which cannot be mitigated with time-invariant RTM parameters. These biases are due to sensor error and modeled temperature errors as discussed in De Lannoy et al. (2013). In a separate exercise (not shown herein), we verified that the σ-values absorb biases in geophysical fields: by re-scaling the soil moisture the RMSD and σ-values are jointly reduced.

Based on these new insights, we currently use globally uniform values of $\sigma_m = 3.5$ K and $\sigma_s = 2.3$ K for the SMOS-based calibration of the RTM parameters that will be needed to generate the forthcoming SMAP L4_SM product. This represents an advance over the mere guesses of $\sigma_m = \sigma_s = 1.0$ K used in earlier work (De Lannoy et al., 2013). The resulting parameter estimates optimally mimic the true relationship between land surface fields and Tb. However, larger σ-values likely increase the error of the Tb simulations, because the model parameters will compensate less for local biases in, for example, soil moisture. These issues will be addressed through model development and bias correction inside the data assimilation system.

4.3. Posterior Tb Simulation Performance

The quality of the estimated MAP parameters can be measured by the skill of the corresponding Tb simulations. Fig. 6 shows the misfit between observed and MAP simulated long-term Tb averages and standard deviations

($RMSD_m$, $RMSD_s$, square root of Eq. 4 and 5) across the 24 observations for the calibration and evaluation period, averaged per vegetation class. Table 2 lists the domain-averaged $RMSD_m$ and $RMSD_s$ values during the calibration period. The prior parameters result in $RMSD_m$=12.9 K and $RMSD_s$=3.2 K. The skill is greatly improved after parameter estimation and very similar for cases P, D and D_σ ($RMSD_m$ around 3 K and $RMSD_s$ around 2.5 K), which is not surprising given that the three different cases generate similar MAP parameter estimates. During the evaluation period, the $RMSD_m$ increases up to 8 K for cropland (Fig. 6b), and the $RMSD_s$ reaches values of 5 K for cropland in the evaluation year (Fig. 6d). Cropland exhibits the largest errors, because of known simulation errors (see above). The larger errors in the evaluation period suggest that Tb simulations for interannually varying agricultural areas will be compromised when using time-invariant RTM parameters and climatological vegetation information.

Table 2 also shows the $RMSD_{\overline{m}}$ and $RMSD_{\overline{s}}$ (square root of Eqs. 6 and 7) for the ensemble mean of the Tb simulations $\overline{m_i(\boldsymbol{\alpha})}$ and $\overline{s_i(\boldsymbol{\alpha})}$. Ensembles are generated by sampling the posterior parameter distribution. The limited uncertainty in the posterior parameters results in comparable ensemble mean and deterministic Tb simulations, so that $RMSD_m \sim RMSD_{\overline{m}}$ and $RMSD_s \sim RMSD_{\overline{s}}$.

4.4. Actual and Expected Tb Similation Errors

If the parameter estimation procedure is statistically consistent, then the actual Tb errors ($RMSD_m$, $RMSD_s$, section 4.3) and the estimated residual Tb errors (σ_m, σ_s, section 4.2) should be of similar magnitude. Indeed, for case D_σ the $RMSD_m$ values during the calibration period (Fig. 6a) and

the σ_m estimates (Fig. 5g) show similar variations across vegetation classes. Likewise, the $RMSD_s$ values (Fig. 6b) and the σ_s estimates (Fig. 5i) show similar variations across vegetation classes. For case D, however, the same is not true. Moreover, Table 2 suggests that the domain-averaged ratio of $RMSD_m$ and $RMEnSp_m$ for the long-term average Tb is 2.9 for case D and 1.0 for case D_σ. Similarly, the domain-averaged ratio of $RMSD_s$ and $RMEnSp_s$ is 2.5 for case D and 1.0 for case D_σ. Adequate results are thus only found for case D_σ by estimating σ_m and σ_s, whereas case D falls short with respect to these metrics. Note that during the evaluation period (not shown), the ratios always exceed 1, because of an increased $RMSD_m$ and $RMSD_s$. Nevertheless, case D_σ is still more consistent than case D.

Table 2 also lists the square root of the mean simulation uncertainty due to posterior parameter error, i.e. $RMEnSp_{m,par}$ and $RMEnSp_{s,par}$ (square root of Eqs. 10 and 11). These metrics are not shown for the prior and PSO cases, because those parameter distributions are not sampled adequately. After parameter estimation, $RMEnSp_{m,par}$=0.3 K and $RMEnSp_{s,par}$=0.2 K for case D, whereas $RMEnSp_{m,par}$=1.0 K and $RMEnSp_{s,par}$=0.4 K for case D_σ. The uncertainty associated with the posterior parameter values is only a small fraction of the total residual error, that is, $RMEnSp_{par} << \sigma$ after parameter estimation. This gives us confidence that the RTM parameters have been estimated reliably and that other errors dominate the residual errors or biases in the long-term Tb averages and standard deviations.

Fig. 7 illustrates that the estimated ensemble residual Tb error standard deviations of case D_σ are consistent with the actual residuals between Tb observations and simulations. Specifically, Fig. 7a shows the SMOS observed

$m_{i,o}$ as dots and the GEOS-5 simulated $m_i(\boldsymbol{\alpha}_{MAP})$ for H-polarized Tb at 6 angles at morning overpasses for cases D and D_σ, as horizontal dashes. The error bars around the Tb simulations reflect twice the lumped uncertainty due to ensemble simulation and observation errors, or $2\sigma_{i,m}$ and $2\sigma_{i,s}$. The plotted results are an average over the entire study domain for the morning overpasses. Fig. 7b shows the same for V-polarized Tb, and Figs. 7c and d provide this information for the long-term Tb standard deviations. The results for the 24 observations derived from evening overpasses are very similar and not shown.

If the uncertainty treatment underlying the parameter estimation is consistent, then 95% of the observations are expected to fall within $2\sigma_m$ or $2\sigma_s$ around both sides of the Tb simulations (assuming Gaussian distributions). The error bars for case D_σ envelop all observations, whereas the error bars for case D enclose less than half of all observations. Fig. 7 also explains the nature of the residual misfit. With the exception of the 57.5°-angle, the Tb simulations $m_i(\boldsymbol{\alpha}_{MAP})$ for morning overpasses consistently underestimate the SMOS-observed $m_{i,o}$ for H-polarization (Fig. 7a) and randomly deviate at V-polarization (Fig. 7b). In contrast, the evening simulations $m_i(\boldsymbol{\alpha}_{MAP})$ slightly overestimate the SMOS-observed $m_{i,o}$ at H-polarization (not shown herein, see De Lannoy et al. (2013)). The SMOS-observed $s_{i,o}$ is always larger than the simulated $s_i(\boldsymbol{\alpha}_{MAP})$. This is probably dominated by observation noise, but could also be attributed to an underestimation of the variability in the Tb simulations. For example, an increase in the RTM parameter h not only compensates for a cold Tb bias but simultaneously also reduces the Tb variability. Fig. 7 thus clearly illustrates why the uncertainty estimates

obtained from case D_σ are superior.

4.5. Convergence and Computational Cost

The effectiveness of the parameter sampling algorithm is also measured by the convergence of the sampled posterior distribution to the target distribution. For DREAM$_{(ZS)}$, the potential scale reduction factor $\sqrt{R}$ by Gelman and Rubin (1992) should be near 1 to inspire confidence that the different MCMC chains have converged to the appropriate target distribution. The target variance is estimated based on the variances within and between chains up to the current iteration. The $\sqrt{R}$ metric measures by which scale the posterior distribution at the current iteration would shrink as the number of MCMC iterations would go to infinity to reach the target distribution. Fig. 8 shows the evolution of $\sqrt{R}$, averaged over all estimated parameters and across the study domain. Initially, the values of $\sqrt{R}$ are rather large (due to random initial sampling) before they settle down and reach values close to 1.

Finally, we report that the derivation of the posterior distributions requires approximately 225 seconds of wall time for a single grid cell using DREAM$_{(ZS)}$ on a state-of-the-art, single-processor computing platform. For global applications that involve $10^5 - 10^6$ grid cells, exploration of the posterior distribution may be too costly. Yet, if our main interest is in the MAP values, and we thus only aim at obtaining a long-term unbiased modeling system, then PSO or DREAM$_{(ZS)}$ are both viable options.

5. Conclusions

Accurate estimates of microwave RTM parameters for large-scale L-band applications are difficult to obtain. The available parameter estimates are generally based on small-scale field experiments and often come without any estimate of their uncertainty. This complicates radiative transfer modeling for both the forward simulation of L-band Tb over land and the retrieval of soil moisture based on Tb observations. Our study presents reliable, large-scale estimates of RTM parameter values and their uncertainty, based on the combination of long-term (climatological) information from satellite observations and land model simulations. These estimated parameter values can serve to limit biases in large-scale Tb simulations and soil moisture retrievals. The estimates of parameter uncertainties help to derive better informed uncertainties for long-term (climatological) Tb simulations and retrievals, and they aid in the development of land surface data assimilation systems. This is a significant advancement compared to the common practice of relying on literature-based parameter values or parameter calibration approaches that do not use Bayesian methods with proper characterization of model and observation errors.

More specifically, the present paper expands earlier research reported in De Lannoy et al. (2013) to derive time-invariant RTM parameters and their uncertainties using observations of the long-term average Tb and the long-term Tb standard deviation obtained from SMOS data. The overall objective is to optimize GEOS-5 Tb simulations prior to sequential assimilation of SMOS or SMAP Tb data, such as planned for the SMAP L4_SM product (Reichle et al., 2012), and to examine the uncertainties involved in the opti-

mization. Per grid cell, 48 observations of the long-term SMOS Tb averages and standard deviations are constructed, i.e. Tb averages and standard deviations for 24 different combinations of 6 incidence angles, 2 polarizations and 2 overpass times. The differences between the observed and GEOS-5 simulated long-term Tb averages and standard deviations are minimized (as opposed to minimizing differences between Tb observations and simulations in the time domain) and used along with the prior parameter information to derive posterior parameter estimates.

The full posterior distribution of RTM parameters is derived using MCMC simulation with the DREAM$_{(ZS)}$ algorithm. To our knowledge, this is the first large-scale application of the DREAM$_{(ZS)}$ algorithm for the estimation of RTM parameters and their underlying uncertainty. The results serve as a benchmark to verify the results from simpler parameter optimization algorithms, such as for example PSO. Such algorithms may be desirable for global-scale operational applications that rely on evolving modeling systems in need of frequent re-calibrations, without a need to fully sample the posterior parameter distribution.

Firstly, we verified that the MAP RTM-parameter values derived from converged posterior distributions with DREAM$_{(ZS)}$ can be approximated by a simpler optimization algorithm (PSO), which corroborates our earlier research (De Lannoy et al., 2013). The MAP parameters will later be used in an RTM that is part of a data assimilation system to ensure minimally biased Tb simulations. Secondly, we obtained reliable estimates of parameter uncertainty with DREAM$_{(ZS)}$, which cannot be obtained with PSO. The relative parameter uncertainties are generally less than 25% of the MAP value

for $< h >$, $< \tau >$ and ω, when including an estimation of the residual (observation and simulation) error standard deviations (σ_m, σ_s) of the long-term Tb averages and standard deviations. This relative parameter uncertainty is substantially reduced compared to the relative uncertainty in literature-based parameter values, which easily exceeds 100%.

The third objective of this paper was to quantify the importance of parameter and other errors on long-term Tb simulations. The MAP parameters reduce the RMSD in long-term Tb averages and standard deviations to 3.4 K and 2.3 K, respectively. This is a reduction by 74% and 30% compared to results with prior parameters. Of this total *posterior* Tb uncertainty, only 1 K and 0.4 K is associated with the posterior uncertainty in the parameter values.

The actual RMSD in the long-term Tb averages and standard deviations is matched by the estimated ensemble residual Tb error standard deviation (σ_m=3.5 K, σ_s=2.3 K, assumed homoscedastic), obtained with DREAM$_{(ZS)}$ case D$_\sigma$. In other words, the joint estimation of RTM parameters and σ_m and σ_s results in a balance between actual and expected errors in Tb simulations, and in statistically adequate parameter values and uncertainty estimates. The prior estimate of 1 K for σ_m and σ_s, used in case P, case D and De Lannoy et al. (2013) was thus too low, except for σ_s over forests which exhibit limited Tb variability due to vegetation attenuation. The largest σ_m-values are found in cropped regions where the lumped residual Tb errors are probably dominated by errors in geophysical fields (e.g. vegetation, soil moisture and temperature) that constitute important inputs to the Tb simulations.

In practice, these findings prompt us to revise the global RTM calibra-

tion of De Lannoy et al. (2013) with improved residual error estimates in preparation for the SMAP L4_SM product. The revised approach ensures parameters that optimally describe the true relationship between land surface fields and Tb, while minimally compensating for potential biases in any of these fields. Furthermore, the findings of the present study indicate that, *after* RTM parameter estimation, the residual climatological uncertainties reside in Tb observation error and GEOS-5 Tb simulation error that is not, or only to a limited extent, related to RTM parameter uncertainty.

In summary, the Bayesian inference of the posterior distribution of the RTM parameters ensures reliable Tb simulations with GEOS-5. The MAP parameter estimates guarantee a Tb assimilation system with limited biases and a realistic connection between Tb and land surface fields such as soil moisture and temperature. Furthermore, the $\text{DREAM}_{(\text{ZS})}$ algorithm reveals the importance of observation error and simulation error that cannot be explained by the RTM parameters. The posterior Tb uncertainties in this study pertain to long-term Tb averages and standard deviations and are thus indicative of biases (i.e. long-term errors). These residual biases will be addressed by model refinement and bias mitigation during the assimilation of satellite-observed Tb data.

Acknowledgement

Gabriëlle De Lannoy and Rolf Reichle were supported by the NASA Soil Moisture Active Passive mission. The authors thank Yann Kerr and Ali Mahmoodi for their help with the SMOS data, Valentijn Pauwels for discussions about PSO and the reviewers for their constructive comments.

References

Balsamo, G., J, F. M., Bélair, S., Deblonde, G., 2006. A global root-zone soil moisture analysis using simulated L-band brightness temperature in preparation for the hydros satellite mission. Journal of Hydrometeorology 7, 1126–1146.

CESBIO, IPSL, INRA, Reading University, Tor Vergata University, 2011. SMOS level 2 processor for soil moisture. Tech. rep., CEBIO, IPLS-Service d'Aeronomie, INRA-EPHYSE, Reading University, Tor Vergata University.

Choudhury, B. J., Schmugge, T. J., Chang, A., Newton, R. W., 1979. Effect of surface roughness on the microwave emission from soils. Journal of Geophysical Research 84 (C9), 5699–5706.

Crosson, W., Limaye, A., Laymon, C., 2005. Parameter sensitivity of soil moisture retrievals from airborne L-band radiometer measurements in SMEX02. IEEE Transactions on Geoscience and Remote Sensing 43, 1517–1528.

De Lannoy, G., Reichle, R., Pauwels, V., 2013. Global calibration of the GEOS-5 L-band microwave radiative transfer model over non-frozen land using SMOS observations. Journal of Hydrometeorology 14, 765–785, http://dx.doi.org/10.1175/JHM-D-12-092.1.

De Lannoy, G. J. M., Houser, P. R., Pauwels, V. R. N., Verhoest, N. E. C., 2006. Assessment of model uncertainty for soil moisture through ensemble verification. Journal of Geophysical Research 111 (D10), D10101.1–18, doi:10.1029/2005JD006367.

de Rosnay, P., Calvet, J., Kerr, Y., Wigneron, J., Lematre, F., Escorihuela, M., Muoz-Sabater, J., Saleh, K., Barri, J., Bouhours, G., Coret, L., Cherel, G., Dedieu, G., Durbe, R., Fritz, N., Froissard, F., Hoedjes, J., Kruszewski, A., Lavenu, F., Suquia, D., Waldteufel, P., 2006. SMOSREX: A long term field campaign experiment for soil moisture and land surface processes remote sensing. Remote Sensing of Environment 102, 377–389, doi:10.1016/j.rse.2006.02.021.

de Rosnay, P., Drusch, M., Boone, A., Balsamo, G., Decharme, B., Harris, P., Kerr, Y., Pellarin, T., Polcher, J., Wigneron, J., 2009. AMMA land surface model intercomparison experiment coupled to the Community Microwave Emission Model: ALMIP-MEM. Journal of Geophysical Research 114 (D05108), 1–18, doi:10.1029/2008JD010724.

Dobson, M., Ulaby, F., Hallikainen, M., El-Rayes, M. A., 1985. Microwace dielectric behavior of wet soil - part II: Dielectric mixing models. IEEE Transactions on Geoscience and Remote Sensing 23, 35–46.

Drusch, M., Holmes, T., de Rosnay, P., Balsamo, G., 2009. Comparing ERA-40-based L-band brightness temperatures with Skylab observations: A calibration/validation study using the Community Microwave Emission Model. Journal of Hydrometeorology 10, 213–226.

Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., Entin, J. K., Goodman, S. D., Jackson, T. J., Johnson, J., Kimball, J., Piepmeier, J. R., Koster, R. D., Martin, N., McDonald, K. C., Moghaddam, M., Moran, S., Reichle, R., Shi, J. C., Spencer, M. W., Thurman, S. W., Tsang, L., Van Zyl, J., 2010. The Soil Moisture Active and Passive (SMAP) mission. Proceedings of the IEEE 98 (5), 704–716.

Gelman, A., Rubin, D., 1992. Inference from iterative simulation using multiple sequences. Stat. Sci. 7, 457–472.

Grant, J., Wigneron, J., Van de Griend, A. A., Kruszewski, A., Sobjaerg, S. S., Skou, N., 2007. A field experiment on microwave forest radiometry: L-band signal behaviour for varying conditions of surface wetness. IEEE Transactions on Geoscience and Remote Sensing 109, 10–19.

Jackson, T., Schmugge, T. J., 1991. Vegetation effects on the microwave emission of soils. Remote Sensing of Environment 36, 203–212.

Kennedy, J., Eberhart, R., 1995. Particle swarm optimization. In: Proc. Int. Conf. Neur. Netw. Piscataway, NJ.

Kerr, Y., Waldteufel, P., Wigneron, J.-P., Delwart, S., Cabot, F., Boutin, J., Escorihuela, M.-J., Font, J., Reul, N., Gruhier, C., Juglea, S., Drinkwater, M., Hahne, A., Martin-Neira, M., Mecklenburg, S., 2010. The SMOS mission: New tool for monitoring key elements of the global water cycle. Proceedings of the IEEE 98 (5), 666–687.

Kerr, Y. H., Njoku, E. G., 1990. A semiempirical model for interpreting

microwave emission from semiarid land surfaces as seen from space. IEEE Transactions on Geoscience and Remote Sensing 28 (3), 384–393.

Kerr, Y. H., Waldteufel, P., Richaume, P., Wigneron, J. P., Ferrazzoli, P., Mahmoodi, A., Al Bitar, A., Cabot, F., Gruhier, C., Juglea, S. E., Leroux, D., Mialon, A., Delwart, S., 2012. The SMOS soil moisture retrieval algorithm. IEEE Transactions on Geoscience and Remote Sensing 50 (5), 1384–1403.

Konings, A., Entekhabi, E., Chan, S. K., Njoku, E. G., 2011. The effect of rtm-parameter on satellite-based soil moisture retrievals. IEEE Transactions on Geoscience and Remote Sensing 49, 2686–2698.

Koster, R. D., Suarez, M. J., Ducharne, A., Stieglitz, M., Kumar, P., 2000. A catchment-based approach to modeling land surface processes in a general circulation model 1. model structure. Journal of Geophysical Research 105 (D20), 24809–24822.

Laloy, E., Vrugt, J., 2012. High-dimensional posterior exploration of hydrologic models using multiple-try DREAM_(ZS) and high-performance computing. Water Resources Research 48 (W01526), doi:10.1029/2011WR010608.

Loveland, T. R., Belward, A. S., 1997. The IGBP-DIS global 1-km land cover data set, DIScover: first results. International Journal of Remote Sensing 65, 1031–1031.

Mironov, V., Dobson, M., Kaupp, V., Komarov, S., Kleshchenko, V., 2004.

Generalized refractive mixing dielectric model for moist soil. IEEE Transactions on Geoscience and Remote Sensing 42, 773–785.

O'Neill, P., Njoku, E., Jackson, T., Chan, S., Bindlish, R., 2012. SMAP Algorithm Theoretical Basis Document: L2 & L3 soil moisture (passive) products. Tech. Rep. SMAP Project, JPL D-66481, Jet Propulsion Laboratory, Pasadena, CA.

Panciera, R., Walker, J. P., Merlin, O., 2009. Improved understanding of soil surface roughness parameterization for L-band passive microwave soil moisture retrieval. IEEE Transactions on Geoscience and Remote Sensing 6 (4), 625–629.

Parinussa, R. M., Meesters, A. G., Liu, Y. Y., Dorigo, W., Wagner, W., de Jeu, R. A., 2011. Error estimates for near-real-time satellite soil moisture as derived from the land parameter retrieval model. IEEE Transactions on Geoscience and Remote Sensing 8 (4), 779–783.

Pellarin, T., Wigneron, J., Calvet, J., Berger, M., Douville, H., Ferrazzoli, P., Kerr, Y. H., Lopez-Baeza, E., Pulliainen, J., Simmods, L. P., Waldteufel, P., 2003. Two-year global simulation of L-band brightness temperatures over land. IEEE Transactions on Geoscience and Remote Sensing 42 (9), 2135–2139.

Reichle, R., Crow, W., Koster, R., Kimball, J., De Lannoy, G., 2012. SMAP Algorithm Theoretical Basis Document: L4 surface and root zone soil moisture product. Tech. Rep. SMAP Project, JPL D-66483, Jet Propulsion Laboratory, Pasadena, CA.

Reichle, R. H., 2012. The MERRA-land data product (version 1.2), GMAO office note No. 3. Tech. rep., NASA Global Modeling and Assimilation Office, available at http://gmao.gsfc.nasa.gov/pubs/office_notes/.

Reichle, R. H., Koster, R. D., De Lannoy, G. J. M., Forman, B. A., Liu, Q., Mahanama, S. P. P., Toure, A., 2011. Assessment and enhancement of MERRA land surface hydrology estimates. Journal of Climate 24, 6322–6338.

Reichle, R. H., Walker, J. P., Koster, R. D., Houser, P. R., 2002. Extended vs. ensemble kalman filtering for land data assimilation. Journal of Hydrometeorology 3, 728–740.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., Woollen, J., 2011. MERRA - NASA's modern-era retrospective analysis for research and applications. Journal of Climate 24 (14), 3624–3648, doi:10.1175/JCLI-D-11-00015.

Sabater, J. M., de Rosnay, P., Balsamo, G., 2011. Sensitivity of L-band NWP forward modelling to soil roughness. International Journal of Remote Sensing iFirst, 1–14, doi:10.1080/01431161.2010.507260.

Vrugt, J., Sadegh, M., 2013. Towards diagnostic model calibration and eval-

uation: Approximate Bayesian Computation. Water Resources Research 49, doi: 10.1002/wrcr.20354.

Vrugt, J., ter Braak, C., Clark, M., Hyman, J., Robinson, B., 2008. Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. Water Resources Research 44 (W00B09), doi:10.1029/2007WR006720.

Vrugt, J. A., ter Braak, C., Diks, C., Robinson, B. A., Hyman, J. M., Higdon, D., 2009. Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. International Journal of Nonlinear Sciences & Numerical Simulation 10 (3), 271–288.

Wang, J. R., Choudhury, B. J., 1981. Remote sensing of soil moisture content over bare field at 1.4 GHz frequency. Journal of Geophysical Research 86, 5277–5282.

Wang, J. R., Schmugge, T. J., 1980. An empirical model for the complex dielectric permittivity of soils as a function of water content. IEEE Transactions on Geoscience and Remote Sensing GE-18 (4), 288–295.

Wöhling, T., Vrugt, J. A., 2011. Multiresponse multilayer vadose zone model calibration using Markov chain Monte Carlo simulation and field water retention data. Water Resources Research 47, W04510.1–19.

Appendix A. Radiative Transfer Model (RTM)

A diagnostic zero-order (tau-omega) microwave RTM is used to simulate L-band Tb at the top of the atmosphere ($\mathrm{Tb}_{TOA,p}$ [K]). The $\mathrm{Tb}_{TOA,p}$ at polarization $p = [H, V]$ (horizontal or vertical) is a combination of (i) soil emission, possibly attenuated by vegetation, (ii) vegetation emission, possibly reflected by the soil, and (iii) atmospheric effects:

$$\begin{aligned} \mathrm{Tb}_{TOV,p} &= \mathrm{T}_s(1-r_p)A_p + \mathrm{T}_c(1-\omega_p)(1-A_p)(1+r_pA_p) \\ &\quad +\mathrm{Tb}_{ad,p}r_pA_p^2 \qquad (A.1) \\ \mathrm{Tb}_{TOA,p} &= \mathrm{Tb}_{au,p} + \exp(-\tau_{atm,p})\mathrm{Tb}_{TOV,p} \qquad (A.2) \end{aligned}$$

where $\mathrm{Tb}_{TOV,p}$ [K] is the top of vegetation Tb, T_s [K] is the surface soil temperature, T_c [K] is the canopy temperature (assumed equal to T_s), $\mathrm{Tb}_{ad,p}$ [K] and $\mathrm{Tb}_{au,p}$ [K] are the downward and upward atmospheric radiation, A_p [-] is the vegetation attenuation, $\exp(-\tau_{atm,p})$ [-] is the atmospheric attenuation, $\tau_{atm,p}$ [-] is the atmospheric optical depth, r_p [-] is the rough surface reflectivity, and ω_p [-] is the scattering albedo. The atmospheric contributions ($\mathrm{Tb}_{ad,p}$, $\mathrm{Tb}_{au,p}$ and $\exp(-\tau_{atm,p})$) are described by Pellarin et al. (2003). The rough surface reflectivity r_p [-] is derived from the smooth surface reflectivity R_p [-] following Choudhury et al. (1979) and Wang and Choudhury (1981):

$$r_p = (Q \;\; R_q + (1-Q)R_p)\exp(-h)\cos^{Nr_p}(\theta) \qquad (A.3)$$

where Q [-] is the polarization mixing ratio and typically set to 0 for L-band (Kerr and Njoku, 1990), θ [rad] is the incidence angle, h [-] is the roughness parameter accounting for dielectric properties that vary at the sub-wavelength scale, Nr_p [-] is the angular dependence, and $q = V$ for $p = H$

and vice versa. The smooth surface reflectivity R_p [-] is given by the Fresnel equations as a function of the dielectric constant, which itself depends on soil moisture, temperature, texture, incidence angle and wavelength. We select the Wang and Schmugge (1980) soil dielectric mixing model for this study. The results with this model are similar to what is obtained with the Mironov et al. (2004) model, and both are in a better agreement with the SMOS data than the Dobson et al. (1985) model. We include the dependence of h on soil moisture (SM [$\mathrm{m^3.m^{-3}}$]) through a stepwise linear expression (adapted from the proposed SMOS soil moisture retrieval algorithm (CESBIO et al., 2011; Kerr et al., 2012)):

$$h = \begin{cases} h_{max} & \text{if } SM <= wt \\ h_{max} + \frac{h_{min} - h_{max}}{poros - wt} \, (SM - wt) & \text{if } wt < SM <= poros \end{cases} \tag{A.4}$$

where $poros$ [$\mathrm{m^3.m^{-3}}$] and wt [$\mathrm{m^3.m^{-3}}$] are the porosity and transition soil moisture, respectively. The latter is modeled as $wt = 0.48.wp + 0.165$ (Wang and Schmugge (1980)) where wp [$\mathrm{m^3.m^{-3}}$] is the wilting point.

The vegetation attenuation A_p [-] is based on the Jackson and Schmugge (1991) vegetation opacity model:

$$A_p = \exp(-\frac{\tau_p}{\cos\theta}), \text{with} \tag{A.5}$$

$$\tau_p = b_p \, VWC = b_p \, LEWT \, LAI \tag{A.6}$$

where τ_p [-] is the nadir vegetation optical depth, which is a function of a vegetation structure parameter b_p [-] and the vegetation water content (VWC) [$\mathrm{kg.m^{-2}}$]. The latter is modeled as the product of LAI [$\mathrm{m^2.m^{-2}}$] and the leaf equivalent water thickness ($LEWT$) [$\mathrm{kg.m^{-2}}$].

Appendix B. Particle Swarm Optimization (PSO)

The PSO algorithm (Kennedy and Eberhart, 1995) is a global search method that uses a dynamic swarm of particles to explore the parameter space. The best position of each individual particle (cognitive aspect) and of the entire swarm (social aspect) are used to guide the particles towards the optimal solution. The iterative swarm search is performed in several independent repetitions to account for sampling variability.

The fitness of each parameter combination in the swarm is measured by an integrated 'cost' or 'objective function' J [-] that measures the distances between the observed and simulated long-term Tb averages ($J_{m,o}$ [-]) and standard deviations ($J_{s,o}$ [-]). To make sure that the estimated parameter values honor the prior information (as used in DREAM$_{(ZS)}$), we also include a penalty term that quantifies deviations of the parameters from their a priori expected values (J_α [-]). This results in the following definition of the objective function to be minimized:

$$\begin{aligned} J = & \left. \sum_{i=1}^{24} \frac{(m_{i,o} - m_i(\boldsymbol{\alpha}))^2}{2\sigma_{i,m}^2} \right\} J_{m,o} \\ & + \left. \sum_{i=1}^{24} \frac{(s_{i,o} - s_i(\boldsymbol{\alpha}))^2}{2\sigma_{i,s}^2} \right\} J_{s,o} \\ & + \left. \sum_{k=1}^{N_\alpha} \frac{(\alpha_{0,k} - \alpha_k)^2}{2\sigma_{\alpha_{0,k}}^2} \right\} J_\alpha \end{aligned} \tag{B.1}$$

where N_α signifies the number of simultaneously estimated parameters. Note that Eq. B.1 is a polished version of Eq. 6 in De Lannoy et al. (2013), which differs in the presentation of the weight factors and constants. This formulation is essentially similar to the definition of the posterior density used in

DREAM$_{(ZS)}$. The main difference is that PSO handles the prior information of the parameters explicitly as penalty term J_α in the objective function, whereas in DREAM$_{(ZS)}$, the prior parameter distribution is handled independently from the likelihood function by application of Bayes law. Both methods should thus find the same "best" parameter values.

List of Tables

List of Figures

Table 1: Parameters selected (X) for calibration in different experiment cases (P, D, D_σ) with an indication of the allowed parameter range ($[\alpha_{min}, \alpha_{max}]$) and the prior estimate for each IGBP vegetation class within the study domain. $\Delta h \equiv h_{max} - h_{min}$, $\Delta b \equiv b_V - b_H$.

		h_{min} [-]	Δh[−]	ω [-]	b_H [-]	Δb [-]	σ_m [K]	σ_s [K]
	Case P (PSO)	X	X	X	X	X	-	-
	Case D ($DREAM_{(ZS)}$)	X	X	X	X	X	-	-
	Case D_σ ($DREAM_{(ZS)}$)	X	X	X	X	X	X	X
	α_{min}	0	0	0	0	-0.15	1E-5	1E-5
	α_{max}	2.0	1.0	0.3	0.7	0.15	60	40
ENF	Evergreen Needleleaf Forest	1.2	0	0.05	0.33	0	1	1
DBF	Deciduous Broadleaf Forest	1	0	0.05	0.33	0	1	1
MXF	Mixed Forest	1.3	0	0.05	0.33	0	1	1
CSH	Closed Shrublands	0.7	0	0.05	0.3	0	1	1
OSH	Open Shrublands	0.7	0	0.05	0.3	0	1	1
WSV	Woody Savannas	0.7	0	0.05	0.3	0	1	1
GRS	Grasslands	0.1	0	0.05	0.2	0	1	1
CRP	Croplands	0.5	0	0.05	0.15	0	1	1
CRN	Crop and Natural Vegetation	0.7	0	0.05	0.15	0	1	1

Table 2: Domain-averaged parameter values and their uncertainty $stdv[.]$ for the prior distributions and the posterior distributions obtained with cases P, D and D_σ. The RMSD values for MAP and ensemble mean Tb simulations (square root of Eqs. 4-7) and the ensemble standard deviations in Tb simulations due to parameter errors (square root of Eqs. 10-11) are averaged across 24 long-term Tb observations and calculated for the calibration period.

	Prior	P	D	D_σ
$< h >$ [-]	0.59	0.72	0.73	0.72
$< \tau >$ [-]	0.35	0.26	0.26	0.25
ω [-]	0.05	0.09	0.09	0.08
σ_m [K]	-	1.00	1.00	3.45
σ_s [K]	-	1.00	1.00	2.26
$stdv[< h >]$ [-]	0.63	-	0.04	0.10
$stdv[< \tau >]$ [-]	0.27	-	0.02	0.04
$stdv[\omega]$ [-]	0.09	-	0.01	0.02
$stdv[\sigma_m]$ [K]	-	-	-	0.78
$stdv[\sigma_s]$ [K]	-	-	-	0.51
$RMSD_m$ [K]	12.90	2.98	2.90	3.41
$RMSD_s$ [K]	3.21	2.57	2.45	2.25
$RMSD_{\overline{m}}$ [K]	-	-	2.94	3.37
$RMSD_{\overline{s}}$ [K]	-	-	2.51	2.42
$RMEnSp_{m,par}$ [K]	-	-	0.29	1.00
$RMEnSp_{s,par}$ [K]	-	-	0.15	0.39

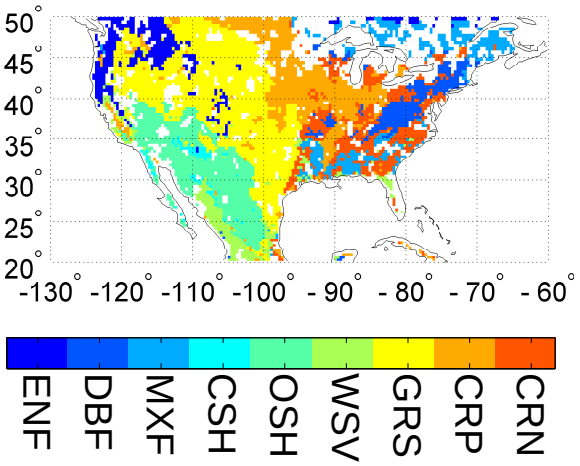

Figure 1: Study domain with indication of the dominant IGBP vegetation classes.

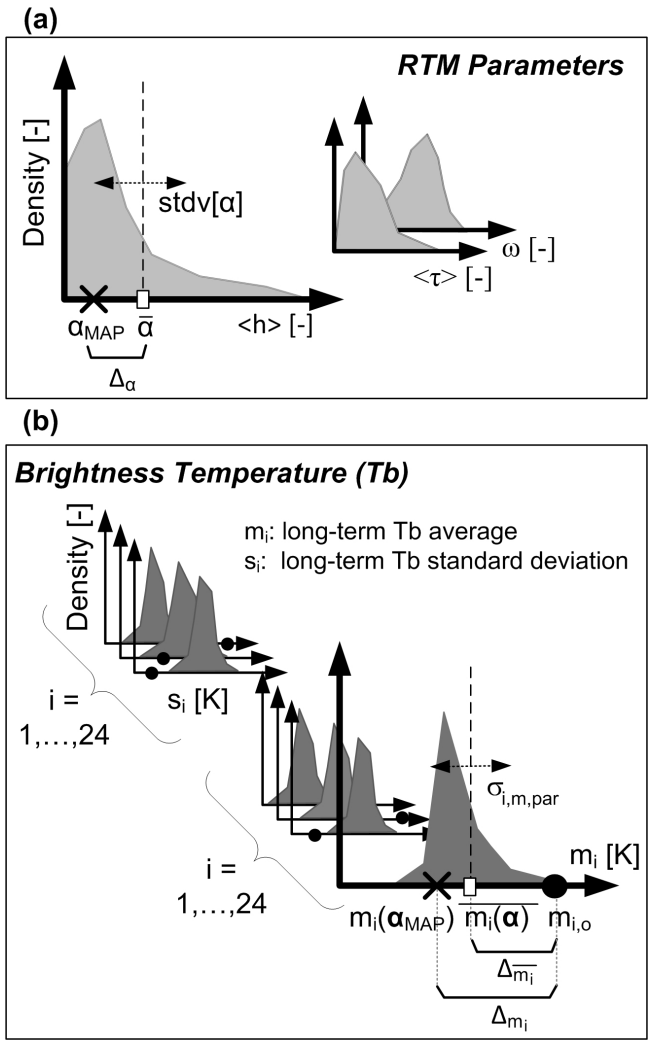

Figure 2: Illustration of marginal distributions for (a) RTM parameters and (b) Tb simulations at a single grid cell. Crosses ($\times$) indicate the MAP estimates, the vertical dashed lines and white squares indicate the ensemble mean posterior estimates, and horizontal dotted arrows indicate one standard deviation uncertainty around the ensemble mean. The performance of the Tb simulations is quantified by comparing either the MAP ($m_i(\boldsymbol{\alpha}_{MAP})$, $s_i(\boldsymbol{\alpha}_{MAP})$) or the ensemble mean ($\overline{m_i(\boldsymbol{\alpha})}$, $\overline{s_i(\boldsymbol{\alpha})}$) simulations against (black dots) 24 observed values $(m_{i,o}, s_{i,o})$ with $i = 1, \ldots, 24$. The differences Δ_{m_i} and $\Delta_{\overline{m}_i}$ contribute to MSD_m (Eq. 4) and $MSD_{\overline{m}}$ (Eq. 6), respectively.

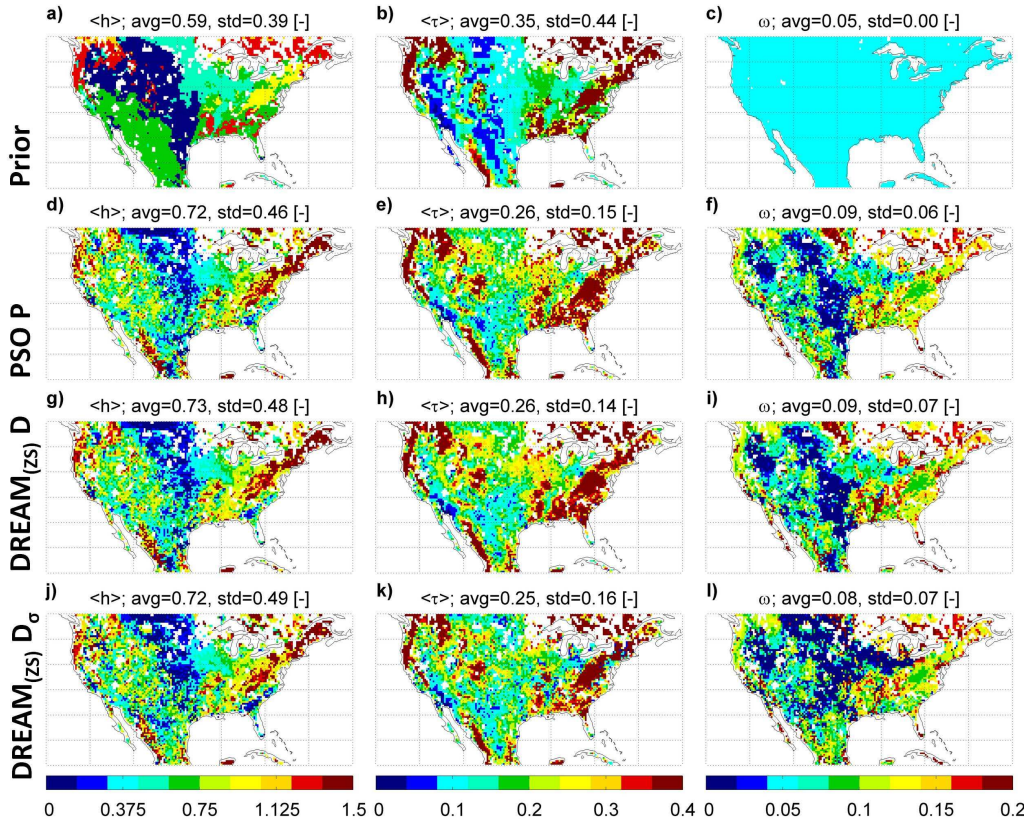

Figure 3: Parameter values for (left) $< h >$, (middle) $< \tau >$, and (right) ω, for the (top row) prior distribution and (second row) case P, (third row) case D and (fourth row) case D_σ.

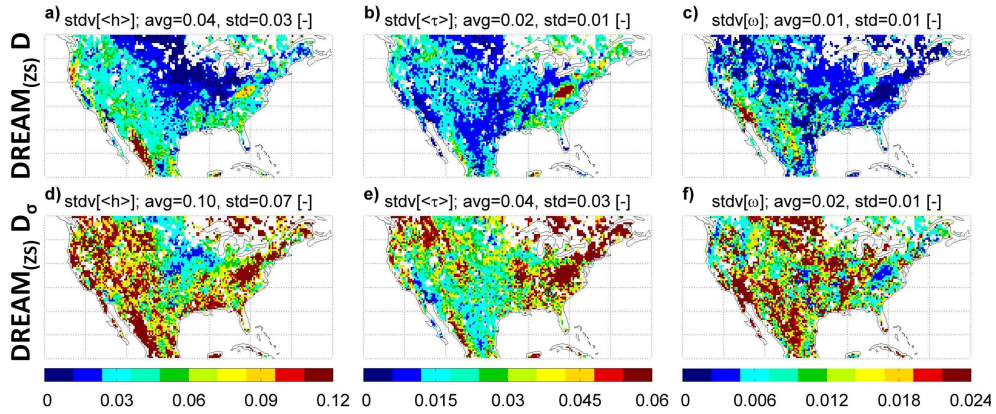

Figure 4: Uncertainty in parameter estimates for (left) $< h >$, (middle) $< \tau >$, and (right) ω, obtained with DREAM$_{(\mathrm{ZS})}$ (top row) case D and (bottom row) case D$_\sigma$. Case D underestimates the posterior parameter uncertainty (see text).

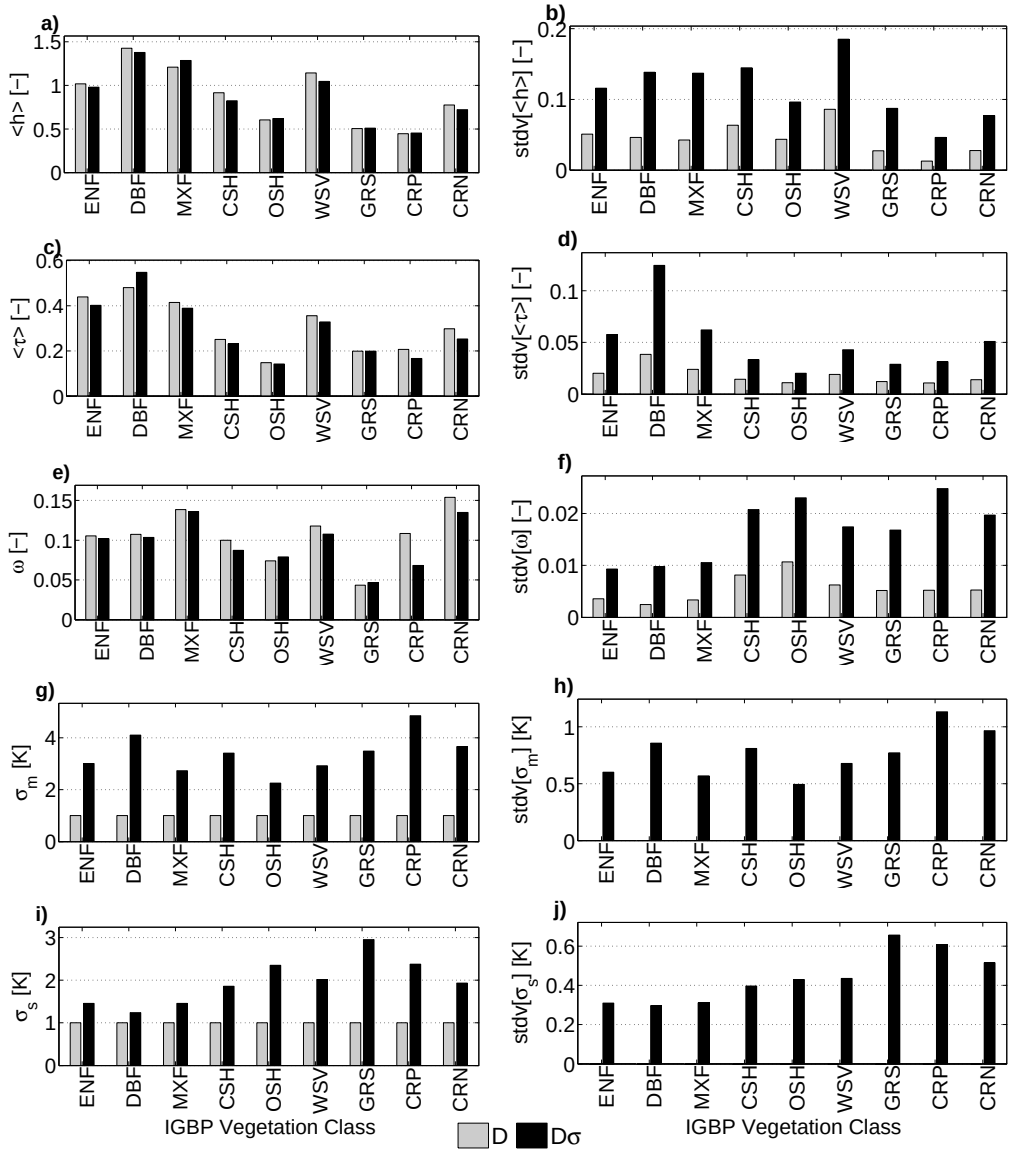

Figure 5: (Left) MAP parameter values and (right) uncertainties aggregated per vegetation class for DREAM$_{(ZS)}$ cases D and D$_\sigma$. Each row represents a different parameter: (a,b) $< h >$, (c,d) $< \tau >$, (e,f) ω, (g,h) σ_m, (i,j) σ_s.

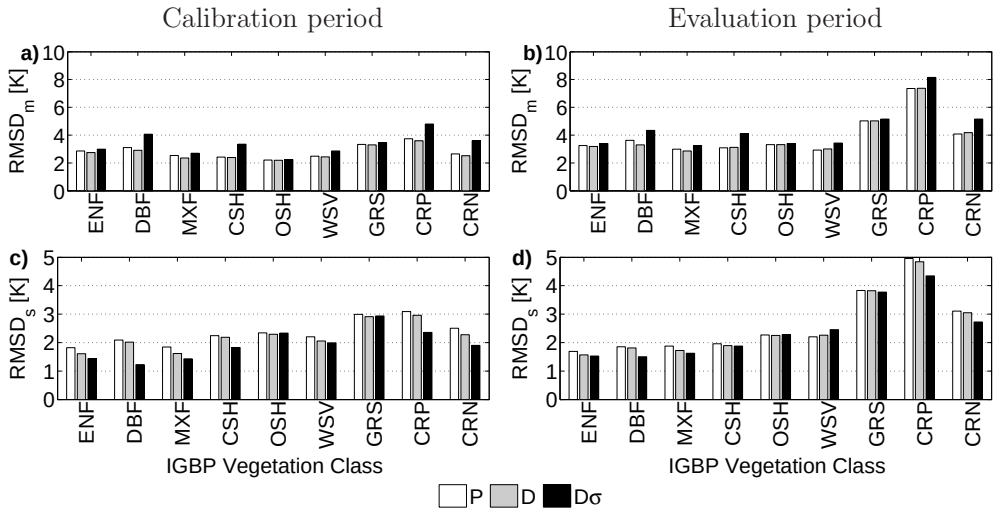

Figure 6: RMSD in long-term Tb (a,b) average and (c,d) standard deviation during the (a,c) calibration period (1 July 2011 - 1 July 2012) and (b,d) evaluation period (1 July 2010 - 1 July 2011), using the MAP parameter values derived from PSO (case P) and $\text{DREAM}_{(\text{ZS})}$ (cases D and D_σ).

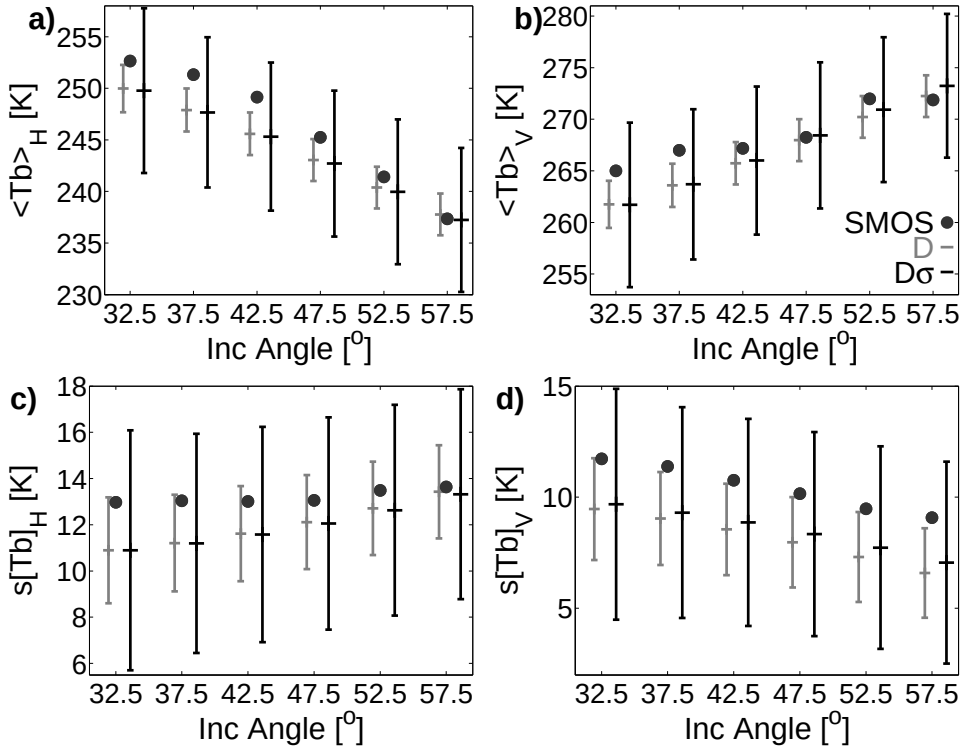

Figure 7: (a,b) Long-term average and (c,d) standard deviation, for (a,c) H- and (b,d) V-polarized Tb from (dots) SMOS observations and (horizontal dashes) simulations averaged over the study domain, during the calibration period (1 July 2011 - 1 July 2012) and only including morning overpasses. The Tb simulations use MAP parameter estimates derived with $\text{DREAM}_{(\text{ZS})}$ (gray) case D and (black) case D_σ, and are shown by a central horizontal dash. The error bars indicate (a,b) $\pm 2\sigma_{i,m}$ and (c,d) $\pm 2\sigma_{i,s}$, i.e twice the lumped standard deviations of the residual simulation and observation errors, and are drawn around the simulated Tb for illustration. For clarity, symbols are slightly offset from the nominal incidence angle.

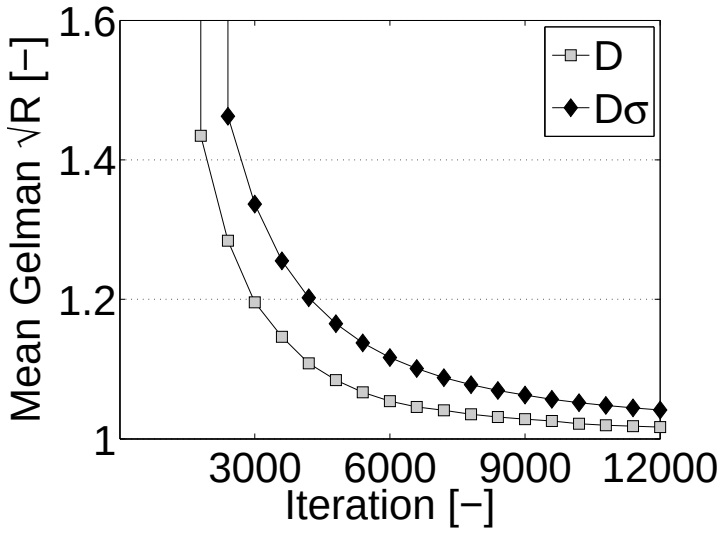

Figure 8: Gelman-Rubin convergence diagnostic $\sqrt{\mathrm{R}}$ for the two DREAM$_{(\mathrm{ZS})}$ MCMC simulation cases. The metric is averaged over all calibrated parameters, and across the study domain.